

JAP20 Rec'd PCT/PTO 22 MAY 2006

## 明細書

## 類似パターン検索装置、類似パターン検索方法、類似パターン検索プログラム、および分画分離装置

## 技術分野

[0001] 本発明は、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索装置、類似パターン検索方法、類似パターン検索プログラム、および分画分離装置に関する。

## 背景技術

[0002] 例えば、フローサイトメトリーは白血球を好中球、リンパ球、単球、好酸球などに短時間で分類することが可能な検査である。フローサイトメトリーによって得られる白血球粒度データは、細胞の成熟度や疾患によってさまざま粒度パターンに分類することができる(非特許文献1参照)。

[0003] この検査は、日常のスクリーニング検査法として多くの施設で導入されているが、分類された数値データのみが利用され、分析装置内部で発生する白血球粒度データを臨床で診断に利用することはまれであった。これは、白血球粒度データが膨大なデータ量であり、外部の情報システムで取り扱うことができなかつた点と、分析生データに対して視覚的な探索方法しか実施されておらず、科学的な手法で検討することが困難であったためである。

[0004] これに対し、本発明者らは、2次元ヒストグラムとして得られる白血球粒度データを用いて、自己組織化マップ(SOM)によりクラスタリングを行う方法を開発した(非特許文献2~4参照)。当該分類方法は、白血球粒度データをデータベースに記録し、データマイニングを適用して特徴的なパターンを抽出することにより、2次元ヒストグラムだけの情報では判断できなかつた分類が可能となる。

[0005] 従来の分類方法は、分析装置内部で各分画の谷間を境界とした分離方法で処理されており、各分画を1つの数値データとして診断に利用する方法が用いられていた。しかし、この方法では、複数のクラスターが近接した分布、たとえば、好中球に属する桿状核球と分葉核球、あるいは、正常細胞と幼弱球の分離ができないという問題が

あつた。

[0006] 非特許文献1:異典之、津田泉、田窪考行、他:自動白血球分類結果の実地診療への反映、HORIBA Technical Reports、No. 20、pp. 23–26、2000。

非特許文献2:片岡浩巳、井沖浩美、小西修、他:白血球粒度のデータマイニング支援システムの構築、日本臨床検査自動化学会誌、Vol27, 4、pp. 583、2002。

非特許文献3:片岡浩巳、井沖浩美、小西修、他:白血球粒度のクラスタリングと3Dビジュアライゼーション、医療情報学 22 (Suppl.)、pp. 209–210、2002。

非特許文献4:井沖浩美、片岡浩巳、川崎由夏、他:白血球粒度データによるアレルギー疾患領域のパターン分類、医療情報学 22 (Suppl.)、pp. 211–212、2002。

## 発明の開示

### 発明が解決しようとする課題

[0007] 本発明は、上記に鑑みてなされたものであって、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを高精度に類似検索して、診断に有用な情報を提供することが可能な類似パターン検索装置、類似パターン検索方法、類似パターン検索プログラム、および分画分離装置を提供することを目的とする。

### 課題を解決するための手段

[0008] 上述した課題を解決し、目的を達成するために、請求項1に係る発明は、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索装置であって、前記複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを選択し、前記複数のパターンに対してクラスタリングを行って作成されたクラスマップを記憶する記憶手段と、前記被検検体のパターンに含まれる成分分画に類似するクラスを前記クラスマップの中から選択する類似パターン検索手段と、を備えたことを特徴とする。

[0009] この請求項1にかかる発明によれば、複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを使用して、複数のパターンに対してクラスタリングを行ってクラスマップを作成し、被検検体のパターンに含まれる成分分画に類似す

るクラスをクラスマップの中から選択して高精度に類似検索を行う。

- [0010] 請求項2にかかる発明は、前記パターンは、1次元または多次元のパターンであることを特徴とする。この請求項2にかかる発明によれば、1次元または多次元のパターンを高精度に類似検索する。
- [0011] 請求項3にかかる発明は、前記パターンは、白血球粒度パターン、蛋白電気泳動波形、または血球ヒストグラムであることを特徴とする。この請求項3にかかる発明によれば、白血球粒度パターン、蛋白電気泳動波形、または血球ヒストグラムを高精度に類似検索する。
- [0012] 請求項4にかかる発明は、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索方法であって、前記複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを選択し、前記複数のパターンにクラスタリングを行ってクラスマップを作成するクラスマップ作成工程と、前記クラスマップ作成工程で作成されたクラスマップを記憶する記憶工程と、前記被検検体のパターンに含まれる成分分画に類似するクラスを前記クラスマップの中から選択する類似パターン検索工程と、を含むことを特徴とする。
- [0013] この請求項4にかかる発明によれば、複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを使用して、複数のパターンに対してクラスタリングを行ってクラスマップを作成し、被検検体のパターンに含まれる成分分画に類似するクラスをクラスマップの中から選択して高精度に類似検索を行う。
- [0014] 請求項5にかかる発明は、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索方法をコンピュータに実行させるためのプログラムであって、前記複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを選択し、前記複数のパターンにクラスタリングを行ってクラスマップを作成するクラスマップ作成工程と、前記クラスマップ作成工程で作成されたクラスマップを記憶する記憶工程と、前記被検検体のパターンに含まれる成分分画に類似するクラスを前記クラスマップの中から選択する類似パターン検索工程と、をコンピュータに実行させることを特徴とする。
- [0015] この請求項5にかかる発明によれば、複数のパターンに各々含まれる複数の成分

分画を特徴づけるモデルパラメータを使用して、複数のパターンに対してクラスタリングを行ってクラスマップを作成し、被検検体のパターンに含まれる成分分画に類似するクラスをクラスマップの中から選択して、高精度に類似検索を行う。

[0016] 請求項6にかかる発明は、複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索装置であって、前記白血球粒度パターンは、複数の細胞成分分画を含み、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段と、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定手段と、前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定手段と、前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリング手段と、前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成手段と、前記2次クラスマップと前記クラス間距離マスターを記憶する記憶手段と、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定手段と、および、前記クラス決定手段で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索手段と、を備えたことを特徴とする。

[0017] この請求項6にかかる発明によれば、自己組織化マップを適用して決定した初期値

を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築する。

[0018] 請求項7にかかる発明は、複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法であって、前記白血球粒度パターンは、複数の細胞成分分画を含み、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング工程と、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定工程と、前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定工程と、前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリング工程と、前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成工程、前記2次クラスマップと前記クラス間距離マスターを記憶する記憶工程と、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定工程と、前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索工程と、を備えたことを特徴とする。

[0019] この請求項7にかかる発明によれば、自己組織化マップを適用して決定した初期値

を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築する。

[0020] 請求項8にかかる発明は、複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法をコンピュータに実行させるプログラムであって、前記白血球粒度パターンは、複数の細胞成分分画を含み、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング工程と、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定工程と、前記各自血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各自血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定工程と、前記各自血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリング工程と、前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成工程と、前記2次クラスマップと前記クラス間距離マスターを記憶する記憶工程と、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定工程と、前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索工程と、をコンピュータに実行させることを特徴とする。

[0021] この請求項8にかかる発明によれば、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築する。

[0022] 請求項9にかかる発明は、複数の細胞成分分画を含む白血球粒度パターンにおいて各細胞成分分画を分離する細胞成分分画分離装置であって、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段と、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分分画の平均値、分散および密度からなる混合分布モデルパラメータを決定するパラメータ決定手段と、および前記各白血球粒度パターンについて、前記混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、各白血球粒度パターンに含まれる各細胞成分の分画を分離する分画分離手段とを備えたことを特徴とする。

[0023] この請求項9にかかる発明によれば、EMアルゴリズムの初期値の決定に自己組織化マップ(SOM)を適用する。

#### 発明の効果

[0024] 本発明(請求項1)にかかる類似パターン検索装置は、複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを使用して、複数のパターンに対してクラスタリングを行ってクラスマップを作成し、被検検体のパターンに含まれる成分分画に類似するクラスをクラスマップの中から選択しているため、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを高精度に類似検索して、診断に有用な情報を提供することができるという効果を奏する。

[0025] 本発明(請求項2)にかかる類似パターン検索装置は、前記パターンとして、1次元または多次元のパターンを使用することとしたので、1次元または多次元の被検検体のパターンと類似度の高いパターンを高精度に類似検索することができるという効果を奏する。

[0026] 本発明(請求項3)にかかる類似パターン検索装置は、前記パターンとして、白血球

粒度パターン、蛋白電気泳動波形、または血球ヒストグラムであることとしたので、白血球粒度パターン、蛋白電気泳動波形、または血球ヒストグラムのパターンと類似度の高いパターンを高精度に類似検索することができるという効果を奏する。

[0027] 本発明(請求項4)にかかる類似パターン検索方法は、複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを使用して、複数のパターンに対してクラスタリングを行ってクラスマップを作成し、被検検体のパターンに含まれる成分分画に類似するクラスをクラスマップの中から選択しているため、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを高精度に類似検索して、診断に有用な情報を提供することができるという効果を奏する。

[0028] 本発明(請求項5)にかかる類似パターン検索プログラムは、複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを使用して、複数のパターンに対してクラスタリングを行ってクラスマップを作成し、被検検体のパターンに含まれる成分分画に類似するクラスをクラスマップの中から選択しているため、複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを高精度に類似検索して、診断に有用な情報を提供することができるという効果を奏する。

[0029] 本発明(請求項6)にかかる類似パターン検索装置は、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築するため、検索対象の類似性を自由に選択できるという効果を奏する。

[0030] 従来は、2次元ヒストグラムの粒度データを直接用いてSOMによりクラスタリングを行っていたため、白血球の個々の成分に注目した部分的な類似性に注目した類似検索を行うことができなかった。本発明によれば、EMアルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能とするものである。

[0031] また、本発明(請求項7)にかかる類似パターン検索方法は、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分

離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築するため、検索対象の類似性を自由に選択できるという効果を奏する。

[0032] また、本発明(請求項8)にかかる類似パターン検索プログラムは、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築するため、検索対象の類似性を自由に選択できるという効果を奏する。

[0033] また、本発明(請求項9)にかかる分画分離装置は、EMアルゴリズムの初期値の決定に自己組織化マップ(SOM)を適用するため、周辺尤度の局所的最大値への収束問題を解決できるという効果を奏する。

#### 図面の簡単な説明

[0034] [図1]図1は、本実施の形態にかかる類似パターン検索装置1の構成を示すブロック図である。

[図2]図2は、本実施の形態にかかる類似パターン検索装置1によって行われる処理のフローチャートである。

[図3]図3は、SOMにより1次クラスタリングを行った結果得られた1次クラスマップの一例を示す図である。

[図4]図4は、もとの粒度データの2次元ヒストグラム(上図)、および、得られた混合分布パラメータを用いて各分画成分を合成して再描画を行ったモデル化された2次元ヒストグラム(下図)である。

[図5]図5は、EMアルゴリズムにより得られた個々の混合分布モデルパラメータをSOMでクラスタリングした結果得られた2次クラスマップの一例を示す図である。

[図6]図6は、好中球領域に分布する桿状核球と分葉核球の分布を示す図である。

[図7]図7は、Class351を基準とした分葉核球の分布の拡大図である。

[図8]図8は、好酸球のClass801を基準とした各クラスの距離をプロットした図である。

[図9]図9は、蛋白電気泳動波形についてSOMにより1次クラスタリングを行った結果

得られた1次クラスマップの一例を示す図である。

[図10]図10は、血球ヒストグラムについてSOMにより1次クラスタリングを行った結果得られた1次クラスマップの一例を示す図である。

[図11]図11は、本発明の一実施例を示す図である。

### 符号の説明

#### [0035] 1 類似パターン検索装置

- 11 1次クラスタリング部
- 12 第1のパラメータ決定部
- 13 第2のパラメータ決定部
- 14 2次クラスタリング部
- 15 クラス間距離マスター作成部
- 16 メモリ
- 17 クラス決定部
- 18 類似パターン検索部

#### 2 分析装置

#### 3 外部入出力装置

### 発明を実施するための最良の形態

[0036] 以下、この発明に係る類似パターン検索装置、類似パターン検索方法、類似パターン検索プログラム、および分画分離装置について、図面を参照しつつ詳細に説明する。なお、この実施形態によりこの発明が限定されるものではない。また、下記実施形態における構成要素には、当業者が容易に想定できるものまたは実質的に同一のものが含まれる。以下の実施形態では、白血球粒度パターンを例示して説明するが、本発明はこれに限られるものではない。

[0037] 本発明者らは、鋭意研究の結果、白血球粒度パターンに含まれる各細胞成分について、EMアルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能にできることを見いだし、この知見に基づいて本発明を完成させるに至った。

[0038] 一般的にEMアルゴリズムは、収束点が初期条件に強く依存し、周辺尤度の局所的最大値を回避できない場合が存在する問題を抱えている。つまり、初期値によっては、低品質の局所解に収束するという現象がある。本発明では、この問題に対し、あらかじめ全体の白血球粒度データをSOMによりクラスタリングした結果をもとに、各クラスの初期値を求め周辺尤度の局所的最大値への収束問題を解決するものである。本発明では、白血球の細胞成分ごと、あるいは、各成分の組み合わせという総合的な視点で高速な類似検索を可能とするアルゴリズムを開発し、診断に有用な情報を提供する。

[0039] 以下に、本発明の実施形態について説明する。図1は、本実施の形態にかかる類似パターン検索装置1の構成を示すブロック図である。本実施の形態にかかる類似パターン検索装置1は、1次クラスタリング部11、第1のパラメータ決定部12、第2のパラメータ決定部13、2次クラスタリング部14、クラス間距離マスター作成部15、メモリ16、クラス決定部17、および類似パターン検索部18を備えている。

[0040] 本発明は、EMアルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能とすることを特徴とする。

[0041] EMアルゴリズムはExpectation step (E-step)とMaximization step (M-step)の2つの処理アルゴリズムから構成され、これらの操作を収束するまで繰り返してパラメータを更新することによって、最尤推定量の極大点を得ることができる。E-stepは、対数尤度の条件付期待値を計算し、M-stepは、条件付期待値を最大化する処理が行われる。

本実施の形態で用いたデータセットと近似モデルは、

データタイプ: 2次元ヒストグラム

モデル: 正規混合モデル

パラメータ: 平均、分散、密度

である。

[0042] 一般的にEMアルゴリズムは、収束点が初期条件に強く依存し、周辺尤度の局所的最大値を回避できない場合が存在する問題を抱えている。つまり、初期値によって

は、低品質の局所解に収束するという現象がある。本発明では、この問題に対し、あらかじめ全体の白血球粒度データをSOMによりクラスタリングした結果をもとに、各クラスの初期値を求め周辺尤度の局所的最大値への収束問題を解決するものである。

- [0043] 分析装置2で測定した白血球粒度の2次元ヒストグラムデータは、類似パターン検索装置1に送信され、メモリ16に格納される。
- [0044] 1次クラスタリング部11は、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成するものである。
- [0045] 第1のパラメータ決定部12は、1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定するものである。
- [0046] 第2のパラメータ決定部13は、実測した各自血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各自血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定するものである。
- [0047] 2次クラスタリング部14は、第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成するものである。なお、ここでは、自己組織化マップを使用することとしたが、K-meanクラスタリング等を使用することにしてもよい。
- [0048] クラス間距離マスター作成部15は、2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するものである。
- [0049] メモリ16は、分析装置2で測定した白血球粒度の2次元ヒストグラムデータ、2次クラスタリング部14で作成した2次クラスマップデータ、クラス間距離マスター作成部15で作成したクラス間距離マスターデータ等のデータを記憶するものである。
- [0050] クラス決定部17は、被検検体の白血球粒度パターンに含まれる各細胞成分分画に

属するクラスを前記2次クラスマップの中から決定するものである。

[0051] 類似パターン検索部18は、クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定するものである。ここでは、類似度を判定する場合に、クラス間距離を使用することとしたが、類似度の評価基準(クラスタ評価基準)はこれに限られるものではなく、クラスタ重心からの距離、クラスタ内距離等を使用することにしてもよい。

[0052] 外部入出力装置2は、ユーザによって入力された各種パラメータや類似パターンの検索条件等を類似パターン検索装置1に送信する。また、類似パターン検索装置1でヒットした類似パターンを画面上に出力する。

[0053] 図2に本実施の形態にかかる類似パターン検索装置1によって行われる処理のフローチャートを示す。なお、以下の説明では、処理の具体例として、自動血球計数装置PENTRA120(堀場製作所)により分析された一般患者検体8,800件のLMNEチャネルの2次元ヒストグラム $128 * 128$ 、8bit／検体のデータを処理するケースに沿って説明を行う。

[0054] (1) 初期値決定のための1次マップの作成  
分析装置2から出力された2次元ヒストグラムデータを近傍8点の平滑化処理を行ったデータに関して、入力層 $128 * 128$ (16,384ニューロン)、競合層 $12 * 12$ (ユニット)のSOMでクラスタリングを行い、得られた144種類のパターンを1次クラスマップとした。SOMの学習パラメータは、近傍距離4、学習率0.3とした。さらに、この1次クラスマップの個々のパターンについて、 $4 * 4$ の16分割した領域を設定し、それぞれの2次元ヒストグラムの重心を求め、その重心を初期値として、EMアルゴリズムにより混合モデルの分離を行った。各分画の分布モデルは正規分布と仮定して計算を行った。さらに、得られた混合分布モデルパラメータ(成分数、各成分の平均値、分散、密度)を人為的に調整し一時的なパラメータを決定した。

[0055] (2) EMアルゴリズムによる混合分布近似  
EMアルゴリズムによる混合分布近似については、「データ学習アルゴリズム」(渡辺

澄夫、共立出版、2001. )およびIgor V. Cadez, Scott Gaffney, Padhraic Smyth : A General Probabilistic Framework for Clustering Individuals and Objects, Knowledge Discovery and Data Mining, pp140–149 , 2000. 等に記載の手法を用いて実施することができる。

[0056] 具体的には、各検体の2次元ヒストグラムデータに対して、1次マップに最も類似したクラスを検索し、そのクラスの混合分布モデルパラメータのパラメータを使ってEMアルゴリズムを実行し粒度成分を分離した。全ての被検粒度データについて同じ処理を実行し個別の混合分布モデルパラメータを計算した。

[0057] (3)混合分布パラメータを入力としたSOMによる2次マップの作成混合分布パラメータを入力としたSOMによる2次マップの作成については、Tom Heskes : Self-organizing maps, vector quantization, and mixture modeling, IEEE Transactions on Neural Networks, 12: pp1299– 1305, 2001. 等に記載の手法を用いて実施することができる。

[0058] 競合層 $30 * 30$ (ユニット)、近傍距離10、学習率0.3のSOMを用いて、X平均値、Y平均値、X共分散行列、Y共分散行列、XY共分散行列、そして、密度の6個で構成された実数の混合分布モデルパラメータを入力層に加えクラスタリングを行った。このクラスタリング結果を2次マップとして類似検索に用いた。この時、各クラス間における全ての組み合わせの類似距離を求めておき、クラス間距離マスターに登録した。

[0059] (4)類似検索  
被検検体の各分画に属するクラスを2次マップから求め、クラス間距離マスターを読み込み、検索の目的にあわせて閾値を決定し、その条件に一致するクラス群を検索した。閾値を可変することにより、検索の類似性の強度を自由に選択できるようにし、さらに、閾値に含まれる領域のクラス群について選言条件で検索することにより類似検索を実現した。各分画の総合的なパターンについての検索を行いたい場合は、それぞれの分画に属するクラスの連言により検索することとした。

[0060] 図3にSOMにより1次クラスタリングを行った結果を示す。12 \* 12の競合層の内部を表示したもので、白血球粒度の全体のパターンを144個にクラスタリングされた結果が得られた。

[0061] 図4の上図は、もとの粒度データの2次元ヒストグラムを示し、+が初期値、×がEMアルゴリズムにより最適尤度の探索が行われた経路と収束点を示す。図4の下図は、得られた混合分布パラメータを用いて各分画成分を合成して再描画を行ったモデル化された2次元ヒストグラムである。

[0062] 図5は、EMアルゴリズムにより得られた個々の混合分布モデルパラメータをSOMでクラスタリングした結果を示す。赤で描画した楕円形の成分は、1成分の細胞の分画を示し、周囲に類似したパターンが配置された結果が得られた。それぞれの細胞群についてさまざまなパターンが存在していることが理解できる。ピンク色1はリンパ球、黄色2は単球、水色3で示す領域は好中球、紫色4は好酸球の分布結果が得られた。文字どおりのLMNEチャンネルの4つの細胞群がクラスタリングされた結果を得た。さらに、リンパ球の下部に分布した白色領域は血小板がマッピングされており、その他の白色部分と各細胞群の境界領域には異常細胞と考えられる分布がマッピングされた。図5と図6に示した細胞群は、左上の角をClass0とし、右下の角をClass899としたラスター方向のシーケンシャル番号で呼ぶことにした。

[0063] 図6は、好中球領域に分布する桿状核球と分葉核球の分布を示す。顕微鏡による目視分類結果よりClass120は桿状核球がどのクラスよりも多いクラスで、Class351は、分葉核球が多い症例のクラスである。黄色31(左)のグラディエーション領域は、核左方移動が著明な桿状核球が一番多く含まれる症例群であったClass120を中心とした類似距離のパターンを色の強度で表現した分布である。また、青色32(右)のグラディエーション領域は、分葉核球が一番多かったClass351を中心としたパターンを示している。

[0064] 図7は、Class351を基準とした分葉核球の分布を拡大した図である。幅広い範囲で類似検索を行いたい場合、赤線で囲んだ領域のクラスを対象に検索し、類似性が強い細胞の検索を行いたい場合は、緑線、あるいは、青線で囲まれた領域のクラスを検索することで検索対象の絞込みが可能となった。

[0065] 図6で示した桿状核球(Class120)と分葉核球(Class351)との視覚的な境界面は、滑らかなグラディエーションで結合されており、類似性の境界が不明瞭であることを示している。これは、桿状核球と分葉核球が同じ好中球に属することから、細胞の分

化度の視点から見た類似性がマップ上でクラスタリングができていることを示唆している。一方、Class351の分葉核球とリンパ球領域との境界はグラディエーションの少ない明瞭な境界面が観察され、これらの細胞群はマップ上で明確に分離できることを示している。図8は、好酸球のClass801を基準とした各クラスの距離をプロットした図である。縦軸は、Class801からの距離、横軸は、距離の少ない順にソートしたクラスである。距離が1以下は同じ好酸球が分布しており、距離の閾値を可変することで、検索対象の類似性を可変できることを示している。また、細胞ごとに階段状の曲線が得られ、好中球の分葉核球と桿状核球が単球で分断されている興味深い結果が得られた。これは、基準となる細胞によってさまざまなパターンとなる傾向があった。

[0066] 白血球の各成分単独、あるいは、各成分を統合した類似性に関して、類似性の尺度を自由に可変できる類似検索システムを構築した。EMアルゴリズムは、事前にSOMによりクラスタリングしたパターンで初期値を決定することにより、正しい収束結果が得られた。また、臨床検査領域のフローサイトメトリーで桿状核球と分葉核球の分離は不可能であったが、本法を用いることで容易に分離することが可能となり、診断や治療に有用な情報を提供するシステムを構築することができた。

[0067] 以上本発明にかかる一実施例について図面を参照して詳述してきたが、具体的な構成例はこの一実施例に限られるものではなく、本発明の要旨を逸脱しない範囲の設計変更等があっても本発明に含まれる。

[0068] 例えば、前述した一実施例においては、類似パターン検索装置1で白血球粒度パターンの類似性を検索することとしたが、本発明はこれに限られるものではなく、一次元の蛋白電気泳動波形や血球ヒストグラム等の被検検体パターンの類似性を検索することもでき、各種の被検検体パターンの類似性を検索することができる。また、被検検体パターンは、上述した白血球粒度パターンのような2次元の情報に限られるものではなく、1次元の情報や多次元の情報(時間軸を含む)についても適用可能である。図9は、類似パターン検索装置1で蛋白電気泳動波形についてSOMにより1次クラスタリングを行った結果得られた1次クラスマップの一例を示す図である。図10は類似パターン検索装置1で血球ヒストグラムについてSOMにより1次クラスタリングを行った結果得られた1次クラスマップの一例を示す図である。

[0069] また、前述した一実施例においては、類似パラメータ検索装置1の機能を実現するためのプログラムを図11に示したコンピュータ読み取り可能な記録媒体60に記録して、この記録媒体60に記録されたプログラムを同図に示したコンピュータ50に読み込ませ、実行することにより各機能を実現してもよい。

[0070] 同図に示したコンピュータ50は、上記プログラムを実行するCPU(Central Processing Unit)51と、キーボード、マウス等の入力装置52と、各種データを記憶するROM(Read Only Memory)53と、演算パラメータ等を記憶するRAM(Random Access Memory)54と、記録媒体60からプログラムを読み取る讀取装置55と、ディスプレイ、プリンタ等の出力装置56とから構成されている。

[0071] CPU51は、讀取装置55を経由して記録媒体60に記録されているプログラムを読み込んだ後、プログラムを実行することにより、前述した機能を実現する。なお、記録媒体60としては、光ディスク、フレキシブルディスク、ハードディスク等が挙げられる。

産業上の利用可能性

[0072] 以上のように、本発明にかかる類似パターン検索装置は、各成分を統合した類似性に関して、類似性の尺度を自由に可変できるため、診断や治療に有用な情報を提供することができる。

## 請求の範囲

[1] 複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索装置であって、  
前記複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを選択し、前記複数のパターンに対してクラスタリングを行って作成されたクラスマップを記憶する記憶手段と、  
前記被検検体のパターンに含まれる成分分画に類似するクラスを前記クラスマップの中から選択する類似パターン検索手段と、  
を備えたことを特徴とする類似パターン検索装置。

[2] 前記パターンは、1次元または多次元のパターンであることを特徴とする請求項1に記載の類似パターン検索装置。

[3] 前記パターンは、白血球粒度パターン、蛋白電気泳動波形、または血球ヒストグラムであることを特徴とする請求項2に記載の類似パターン検索装置。

[4] 複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索方法であって、  
前記複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを選択し、前記複数のパターンにクラスタリングを行ってクラスマップを作成するクラスマップ作成工程と、  
前記クラスマップ作成工程で作成されたクラスマップを記憶する記憶工程と、  
前記被検検体のパターンに含まれる成分分画に類似するクラスを前記クラスマップの中から選択する類似パターン検索工程と、  
を含むことを特徴とする類似パターン検索方法。

[5] 複数のパターンを含む集団の中から、被検検体のパターンと類似性の高いパターンを検索する類似パターン検索方法をコンピュータに実行させるためのプログラムであって、  
前記複数のパターンに各々含まれる複数の成分分画を特徴づけるモデルパラメータを選択し、前記複数のパターンにクラスタリングを行ってクラスマップを作成するクラスマップ作成工程と、

前記クラスマップ作成工程で作成されたクラスマップを記憶する記憶工程と、  
前記被検検体のパターンに含まれる成分分画に類似するクラスを前記クラスマップ  
の中から選択する類似パターン検索工程と、  
をコンピュータに実行させることを特徴とする類似パターン検索プログラム。

[6] 複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パター  
ンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索  
装置であって、  
前記白血球粒度パターンは、複数の細胞成分分画を含み、  
実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用して  
クラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段と、  
前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEM  
アルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、な  
らびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラ  
メータを決定する第1のパラメータ決定手段と、  
前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初  
期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含ま  
れる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる  
第2の混合分布モデルパラメータを決定する第2のパラメータ決定手段と、  
前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自  
己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラ  
スタリング手段と、  
前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を  
計算し、クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マス  
ターを作成するクラス間距離マスター作成手段と、  
前記2次クラスマップと前記クラス間距離マスターを記憶する記憶手段と、  
被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記  
2次クラスマップの中から決定するクラス決定手段と、  
前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラ

スを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索手段と、  
を備えたことを特徴とする類似パターン検索装置。

[7] 複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法であって、

前記白血球粒度パターンは、複数の細胞成分分画を含み、  
実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用して  
クラスタリングを行い、1次クラスマップを作成する1次クラスタリング工程と、

前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEM  
アルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、な  
らびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラ  
メータを決定する第1のパラメータ決定工程と、

前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初  
期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含ま  
れる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる  
第2の混合分布モデルパラメータを決定する第2のパラメータ決定工程と、

前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自  
己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラ  
スタリング工程と、

前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を  
計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マ  
スターを作成するクラス間距離マスター作成工程と、

前記2次クラスマップと前記クラス間距離マスターを記憶する記憶工程と、  
被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記  
2次クラスマップの中から決定するクラス決定工程と、

前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラ

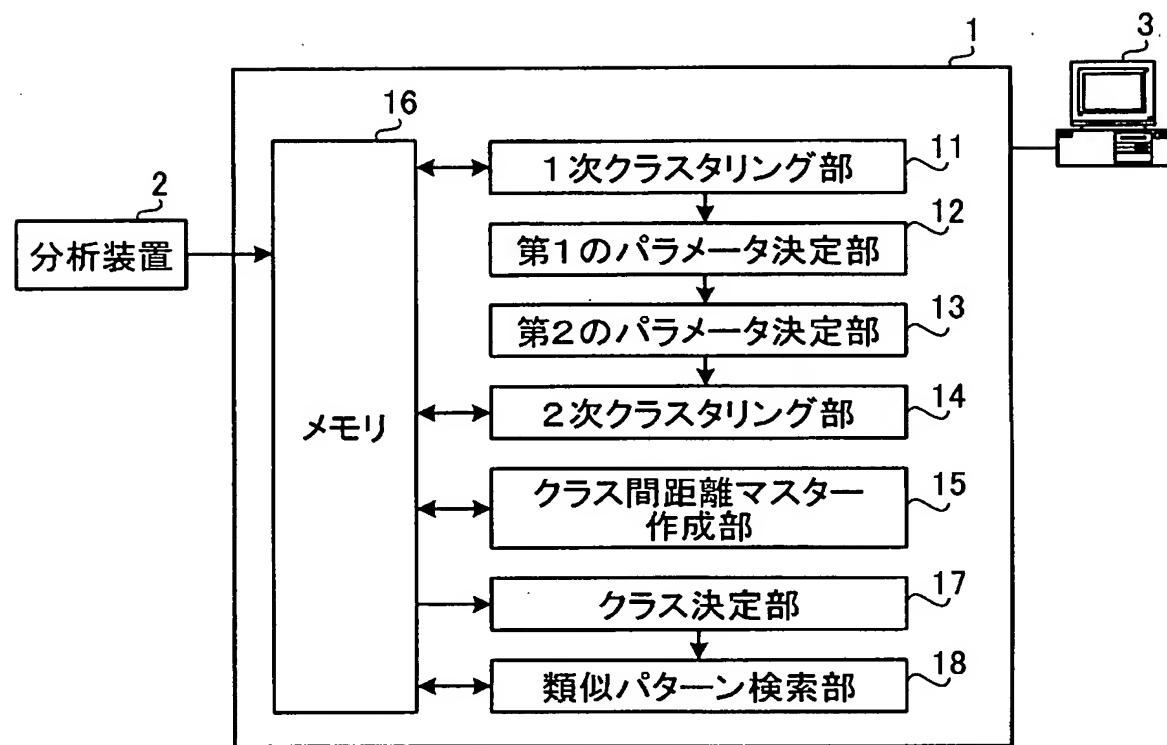
スを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索工程と、  
を含むことを特徴とする類似パターン検索方法。

[8] 複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法をコンピュータに実行させるプログラムであって、  
前記白血球粒度パターンは、複数の細胞成分分画を含み、  
実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング工程と、  
前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定工程と、  
前記各自白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各自白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定工程と、  
前記各自白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリング工程と、  
前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成工程と、  
前記2次クラスマップと前記クラス間距離マスターを記憶する記憶工程と、  
被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定工程と、  
前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラ

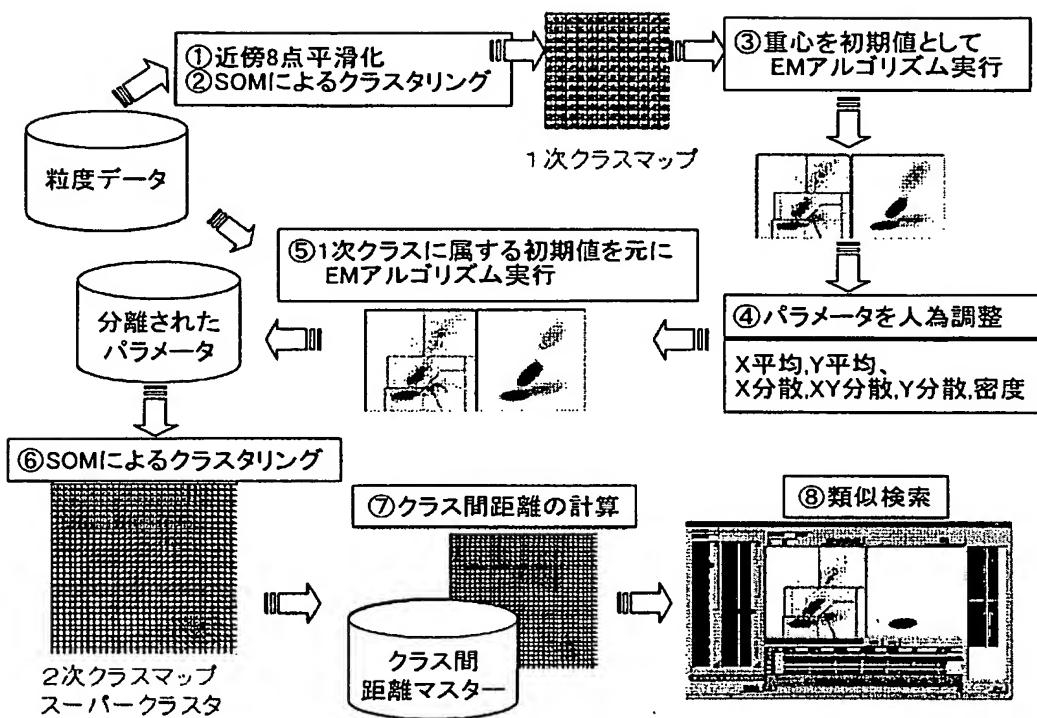
スを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索工程と、  
をコンピュータに実行させることを特徴とする類似パターン検索プログラム。

[9] 複数の細胞成分分画を含む白血球粒度パターンにおいて各細胞成分分画を分離する細胞成分分画分離装置であつて、  
実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段と、  
前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分分画の平均値、分散および密度からなる混合分布モデルパラメータを決定するパラメータ決定手段と、  
前記各白血球粒度パターンについて、前記混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、各白血球粒度パターンに含まれる各細胞成分の分画を分離する分画分離手段と、  
を備えたことを特徴とする分画分離装置。

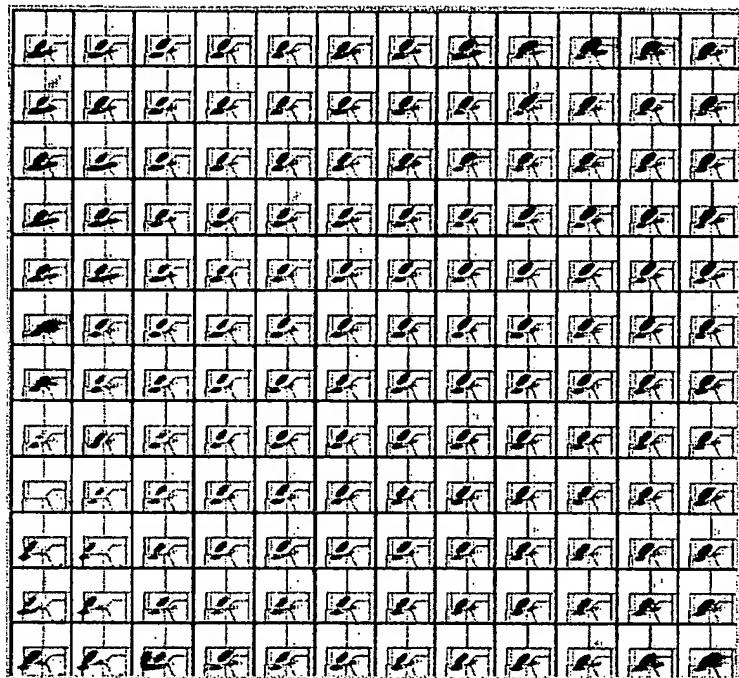
[図1]



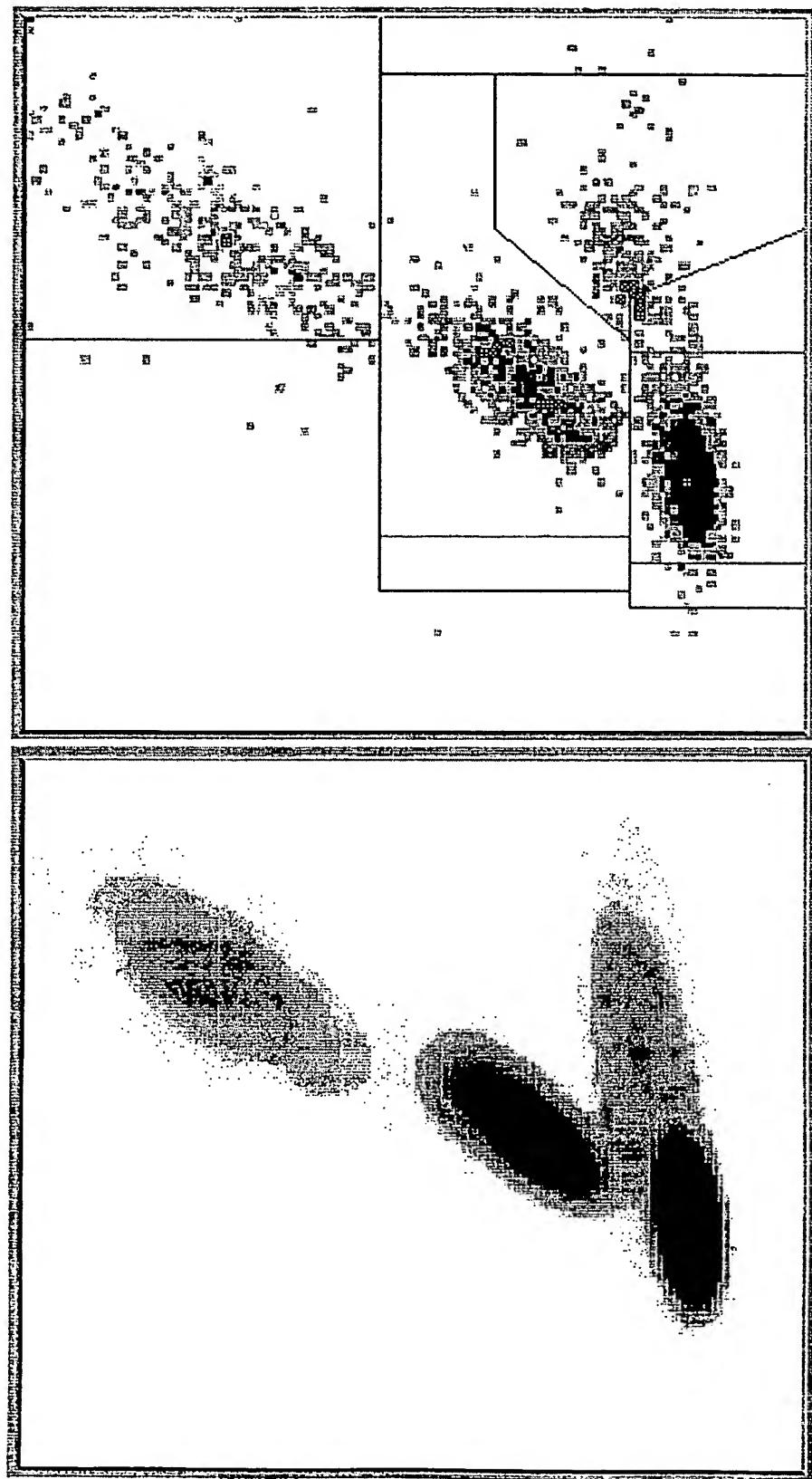
[図2]



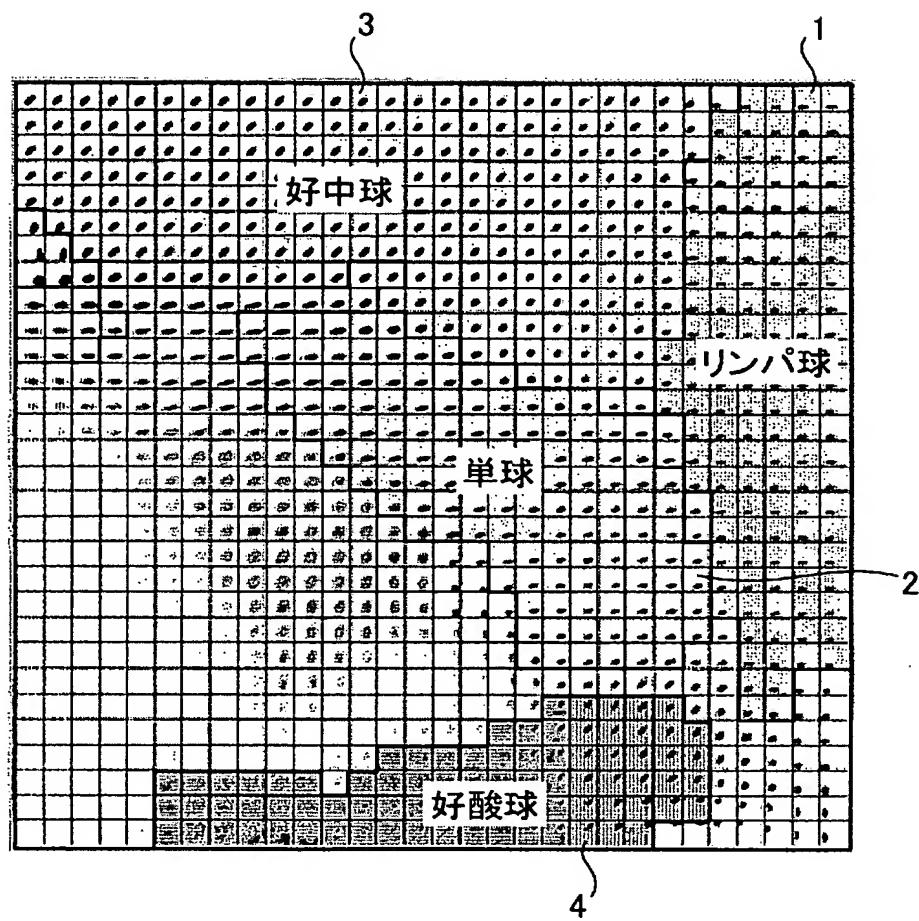
[図3]



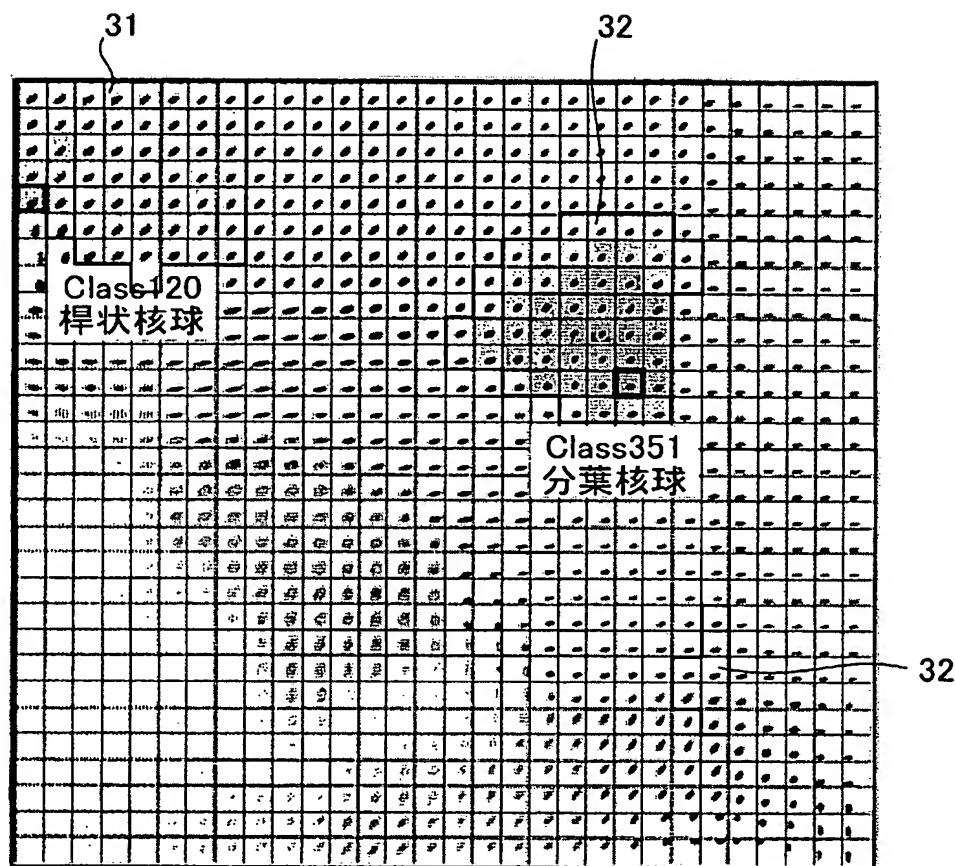
[図4]



[図5]

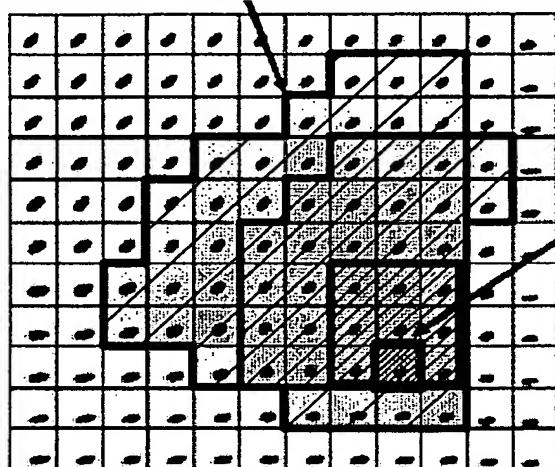


[図6]



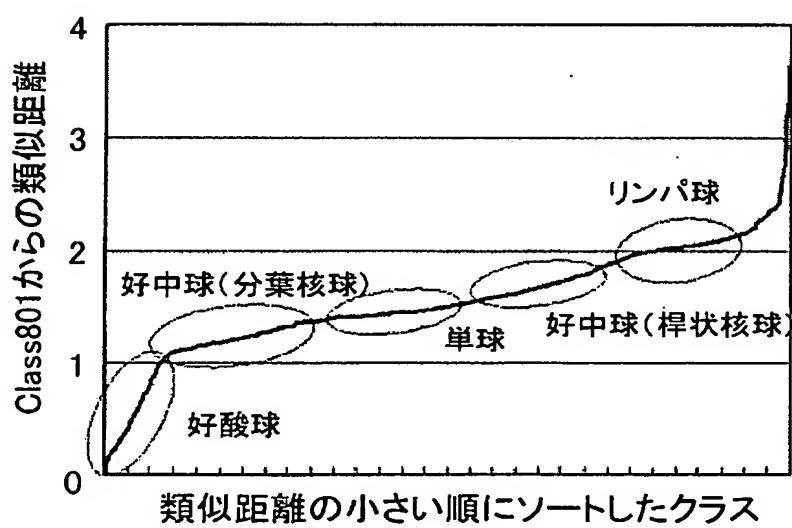
[図7]

Class351を基準とした広い類似距離範囲

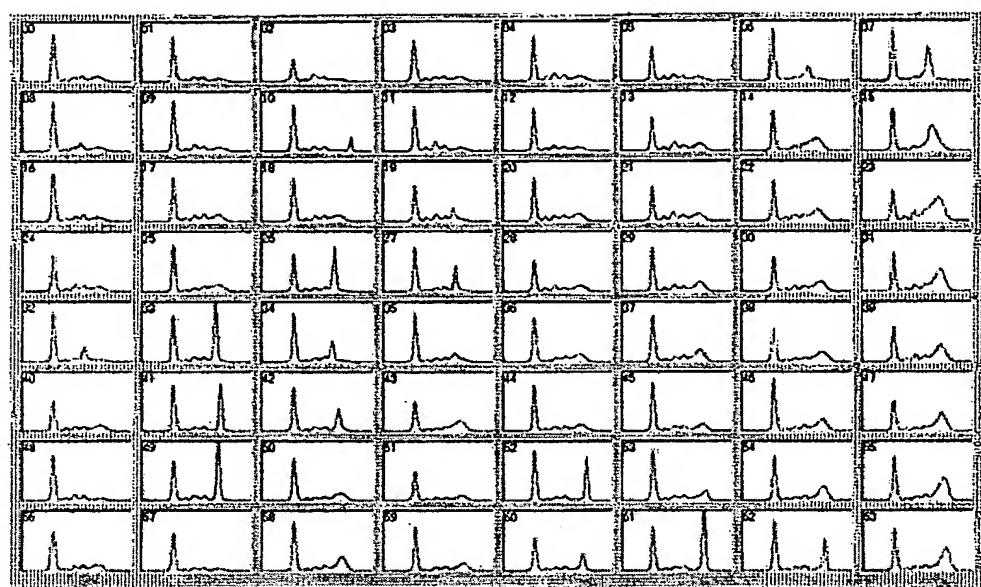
Class351を基準とした  
狭い類似距離範囲

※各クラスの選言により検索の類似性を可変

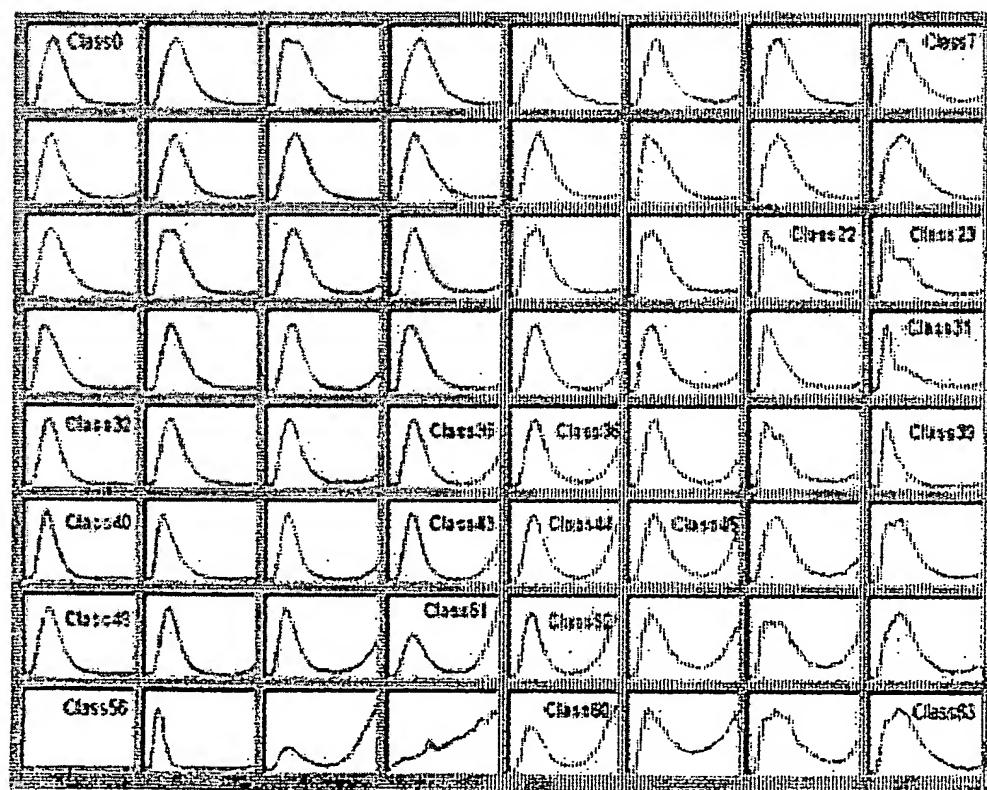
[図8]



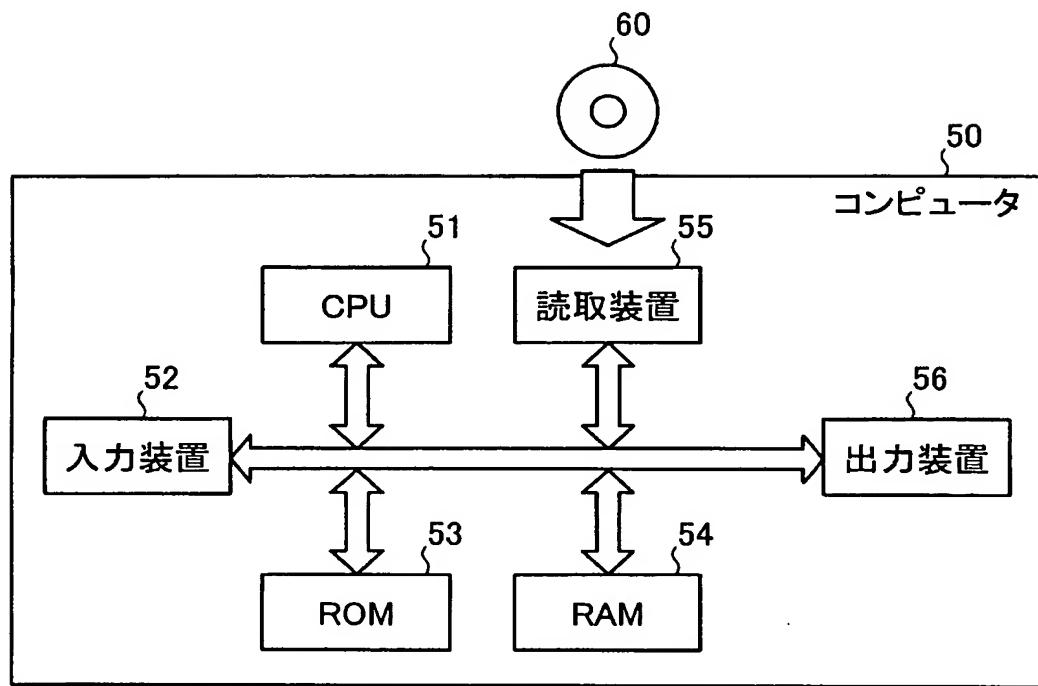
[図9]



[図10]



[図11]



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**